

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36682 A2

(51) International Patent Classification⁷: C12Q 1/68

(74) Agent: FIELD, Bret, E.; Bozicevic, Field & Francis LLP,
Suite 200, 200 Middlefield Road, Menlo Park, CA 94025
(US).

(21) International Application Number: PCT/US00/31562

(22) International Filing Date:
15 November 2000 (15.11.2000)

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/440,829 15 November 1999 (15.11.1999) US

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant: CLONTECH LABORATORIES, INC.
[US/US]; 1020 East Meadow Drive, Palo Alto, CA 94303
(US).

Published:

— Without international search report and to be republished
upon receipt of that report.

(72) Inventors: CHENCHIK, Alex; 670 San Antonio Road
#30, Palo Alto, CA 94306 (US). MUNISHKIN, Alexan-
der; 109 Castillion Terrace, Santa Cruz, CA 95060 (US).
SIMONENKO, Peter; 712 Sierra Vista, Mountain View,
CA 94043 (US).

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.



WO 01/36682 A2

(54) Title: LONG OLIGONUCLEOTIDE ARRAYS

(57) Abstract: Long oligonucleotide arrays, as well as methods for their preparation and use in hybridization assays, are provided. The subject arrays are characterized in that at least a portion of the probes of the array, and usually all of the probes of the array, are long oligonucleotides, e.g. oligonucleotides having a length of from about 50 to 120 nt. Each long oligonucleotide probe on the array is preferably chosen to exhibit substantially the same high target binding efficiency and substantially the same low non-specific binding under conditions in which the array is employed. The subject arrays find use in a number of different applications, e.g. differential gene expression analysis.

LONG OLIGONUCLEOTIDE ARRAYS

INTRODUCTION

5

Technical Field

The field of this invention is nucleic acid arrays.

Background of the Invention

10 Nucleic acid arrays have become an increasingly important tool in the biotechnology industry and related fields. Nucleic acid arrays, in which a plurality of nucleic acids are deposited onto a solid support surface in the form of an array or pattern, find use in a variety of applications, including drug screening, nucleic acid sequencing, mutation analysis, and the like. One important use of nucleic acid arrays is in the analysis of differential gene expression, where the expression of genes in different cells, normally a cell of interest and a control, is compared and any discrepancies in expression are identified. In such assays, the presence of discrepancies indicates a difference in the classes of genes expressed in the cells being compared.

15 In methods of differential gene expression, arrays find use by serving as a substrate to which is bound nucleic acid "probe" fragments. One then obtains "targets" from at least two different cellular sources which are to be compared, e.g. analogous cells, tissues or organs of a healthy and diseased organism. The targets are then hybridized to the immobilized set of nucleic acid "probe" fragments. Differences between the resultant hybridization patterns are then detected and related to differences in gene expression in the two sources.

20 A number of different physical parameters of the array which is used in such assays can have a significant effect on the results that are obtained from the assay. One physical parameter of nucleic acid arrays that can exert a significant influence over the nature of the results which are obtained from the array is probe size, i.e. the length of the individual probe nucleic acids stably associated with the surface of the solid support in the array. There are generally two different types of arrays currently finding use--(1) cDNA arrays, in which either full length or partial cDNAs are

employed as probes; and (2) oligonucleotide arrays, in which probes of from about 8 to 25 nucleotides are employed.

In currently used cDNA arrays, the double stranded cDNAs which may be substantially full length or partial fragments thereof are stably associated with the surface of a solid support, e.g. nylon membrane. Advantages of cDNA arrays include high sensitivity, which feature stems from the high efficiency of binding of the cDNA probe to its target and the stringent hybridization and washing conditions that may be employed with such arrays. Disadvantages of cDNA arrays include difficulties in large scale production of such arrays, low reproducibility of such arrays, and the like.

The other current alternative, oligonucleotide arrays, employs oligonucleotide probes in which each probe ranges from about 8 to 35, usually 20 to 35 nucleotides in length. While such arrays are more amenable to large scale production, they suffer from disadvantages as well. One significant disadvantage for such arrays is their lower sensitivity for target nucleic acids, as compared to cDNA arrays. Another disadvantage is the wide variation in hybridization efficiency of different probes for the same target in a given protocol, which feature requires the use of multiple oligonucleotide probes for the same target, which redundancy adds significantly to the cost of producing such arrays.

As such, there is a continued interest in the development of new array formats. Of particular interest would be the development of array format which combined the high sensitivity of cDNA arrays with the high throughput manufacturability of oligonucleotide arrays, where the format would not suffer from the disadvantages experienced with cDNA and oligonucleotide arrays, as described above.

SUMMARY OF THE INVENTION

Long oligonucleotide arrays, as well as methods for their preparation and use in hybridization assays, are provided. The subject arrays are characterized in that at least a portion of the probes of the array, and usually all of the probes of the array, are long oligonucleotides, e.g. oligonucleotides having a length of from about 50 to 120 nt. Each long oligonucleotide probe on the array is preferably chosen to exhibit high target binding efficiency and low non-specific binding under conditions in which the array is employed, e.g., stringent hybridization conditions. In many embodiments, the specific probe oligonucleotides are chosen so that they have substantially the same hybridization efficiency to their respective targets. The subject arrays find use in a number of different applications, e.g. differential gene expression analysis.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 provides a graphical representation of the hybridization efficiency of different length oligonucleotides.

DEFINITIONS

The term "nucleic acid" as used herein means a polymer composed of nucleotides, e.g., naturally occurring deoxyribonucleotides or ribonucleotides, as well as synthetic mimetics thereof which are also capable of participating in sequence specific, Watson-Crick type hybridization reactions, such as is found in peptide nucleic acids, etc.

The terms "ribonucleic acid" and "RNA" as used herein mean a polymer composed of ribonucleotides.

The terms "deoxyribonucleic acid" and "DNA" as used herein mean a polymer composed of deoxyribonucleotides.

The term "short oligonucleotide" as used herein denotes single stranded nucleotide multimers of from about 8 to 50 nucleotides in length, i.e. 8 to 50 mers.

The term "long oligonucleotide" as used herein denotes single stranded nucleotide multimers of from about 50 to 150, usually from about 50 to 120, nucleotides in length, e.g. a 50 to 150 mer, 50 to 120 mer, etc.

The term "polynucleotide" as used herein refers to single or double stranded polymer composed of nucleotide monomers of greater than about 150 nucleotides in length up to about 5000 nucleotides in length.

The term "oligonucleotide probe composition" refers to the nucleic acid composition that makes up each of the probes spots on the array that correspond to a target nucleic acid. Thus, oligonucleotide probe compositions of the subject arrays are nucleic acid compositions of a plurality of long oligonucleotides, where the composition may be homogenous or heterogenous with respect to the long oligonucleotides that make up the probe composition, i.e., each of the long oligonucleotides in the probe composition may have the same sequence such that they are identical or each of the probe compositions may be made up of two or more different long oligonucleotides that differ from each other in terms of sequence.

The term "target nucleic acid" means a nucleic acid for which there is one or more corresponding oligonucleotide probe compositions, i.e., probe oligonucleotide spots, present on the array. The target nucleic acid may be represented by one or more different oligonucleotide probe compositions on the array. The target nucleic acid is a nucleic acid of interest in a sample being tested with the array, where by "of interest" is meant that the presence or absence of target in the

sample provides useful information, e.g., unique and defining characteristics, about the genetic profile of the cell(s) from which the sample is prepared. As such, target nucleic acids are not housekeeping genes or other types of genes which are present in a number of diverse cell types and therefore the presence or absence of which does not provide characterizing information about a particular cell's genetic profile.

The terms "background" or "background signal intensity" refers to hybridization signals resulting from non-specific binding of labeled target to the substrate component of the array. Background signals may also be produced by intrinsic fluorescence of the array components themselves. A single background signal can be calculated for the entire array, or a different background signal may be calculated for each target nucleic acid.

The term "non-specific hybridization" refers to the non specific binding or hybridization of a target nucleic acid to a nucleic acid present on the array surface, e.g., a long oligonucleotide probe of a probe spot on the array surface, a nucleic acid of a control spot on the array surface, and the like, where the target and the probe are not substantially complementary.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Long oligonucleotide arrays, as well as methods for their preparation and use in hybridization assays, are provided. The subject arrays are characterized in that at least a portion or fraction, usually a majority of or substantially all of the probes of the array, and usually all of the probes of the array, are long oligonucleotides, e.g., oligonucleotides having a length of from about 50 to 120 nt. Each long oligonucleotide probe on the array is preferably chosen to exhibit high target binding efficiency and low non-specific hybridization under conditions in which the array is employed, e.g., stringent conditions. In certain embodiments, the arrays are further characterized in that each of the distinct probes on the array has substantially the same hybridization efficiency for its respective target. The subject arrays find particular use in gene expression assays. In further describing the subject invention, the arrays will first be described in general terms. Next, methods for their preparation are described. Following this description, a review of representative applications in which the subject arrays may be employed is provided.

Before the subject invention is described further, it is to be understood that the invention is not limited to the particular embodiments of the invention described below, as variations of the particular embodiments may be made and still fall within the scope of the appended claims. It is also to be understood that the terminology employed is for the purpose of describing particular

embodiments, and is not intended to be limiting. Instead, the scope of the present invention will be established by the appended claims.

5 In this specification and the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this invention belongs.

ARRAYS OF THE SUBJECT INVENTION-GENERAL DESCRIPTION

10 The arrays of the subject invention have a plurality of probe spots stably associated with a surface of a solid support. A feature of the subject arrays is that at least a portion of the probe spots, and preferably substantially all of the probe spots on the array are probe oligonucleotide spots, where each probe oligonucleotide spot on the array comprises an oligonucleotide probe composition made up of a plurality of long oligonucleotides of known identity, usually of known
15 sequence, as described in greater detail below.

Probe Spots of the Arrays

As mentioned above, a feature of the subject invention is the nature of the probe spots, i.e., that at least a portion of, and usually substantially all of, the probe spots on the array are made up of
20 probe nucleic acid compositions of long oligonucleotides. Each probe spot on the surface of the substrate is made up of long oligonucleotide probes, where the spot may be homogeneous with respect to the nature of the long oligonucleotide probes present therein or heterogenous, e.g., as described in U.S. Patent Application Serial No. 09/417,268, the disclosure of which is herein incorporated by reference. A feature of the oligonucleotide probe compositions is that the probe
25 compositions are made up of long oligonucleotides. As such, the oligonucleotide probes of the probe compositions range in length from about 50 to 150, typically from about 50 to 120 nt and more usually from about 60 to 100 nt, where in many preferred embodiments the probes range in length from about 65 to 85 nt. In other words, in each probe spot or composition, the length of each of the individual probes that make up the probe spot or composition falls within the above ranges.

30 In addition to the above length characteristics, the long oligonucleotide probes that make up the probe spots in the above are typically characterized by one or more of the following features in many preferred embodiments of the subject invention. One further characterization of the long oligonucleotides probes that make up the subject arrays is that their sequence is chosen to provide for high binding efficiency to their complementary target under stringent conditions. Binding

efficiency refers to the ability of the probe to bind to its target under the hybridization conditions in which the array is used. Put another way, binding efficiency refers to the duplex yield obtainable with a given probe and its target after performing a hybridization experiment. In many embodiments, the probes present on the array surface that exhibit high binding efficiency have a binding efficiency for their target of 0.1%, usually at least 0.5 % and more usually at least 2%.

Furthermore, the sequence of the long oligonucleotide probes is chosen to provide for low non-specific hybridization or non-specific binding, i.e., unwanted cross-hybridization, to target nucleic acids for which the probes are not substantially complementary under stringent conditions. A given target is considered to be substantially non-complementary to a given probe in the target has homology to the probe of less than 60%, more commonly less than 50% and most commonly less than 40%, as determined using the BLAST program with default settings. In certain embodiments, oligonucleotide probes having low non-specific hybridization characteristics and finding use in the subject arrays are those in which their relative ability to hybridize to non-complementary nucleic acids, i.e., other targets for which they are not substantially complementary, is less 10 %, usually less than 5 % and preferably less than 1 % of their ability to bind to their complementary target. For example, in a side-by-side hybridization assay, probes having low non-specific hybridization characteristics are those which generate a positive signal, if any, when contacted with a target composition that does not include a complementary target for the probe, that is less than about 10%, usually less than about 3% and more usually less than about 1% of the signal that is generated by the same probe when it is contacted with a target composition that includes a complementary target.

In addition, the long oligonucleotides of a given spot are chosen so that each long oligonucleotide probe present on the array, or at least its target specific sequence, is not homologous with any other distinct unique long oligonucleotide present on the array, i.e., any other oligonucleotide probe on the array with a different base sequence. In other words, each distinct oligonucleotide of a probe composition corresponding to a first target does not cross-hybridize under stringent conditions (as defined below) with, or have the same sequence as, any other distinct unique oligonucleotide of any probe composition corresponding to a different target, i.e., an oligonucleotide of any other oligonucleotide probe composition that is represented on the array. As such, the sense or anti-sense nucleotide sequence of each unique oligonucleotide of a probe composition will have less than 90% homology, usually less than 70% homology, and more usually less than 50% homology with any other different oligonucleotide of a probe composition corresponding to a different target of the array, where homology is determined by sequence analysis comparison using the FASTA program using default settings. The sequence of unique

The oligonucleotides of each probe composition, or at least the portion of these oligonucleotides that is complementary to their intended targets, i.e., their target specific sequences, are further characterized as follows. First, they have a GC content of from about 35 % to 80%, usually between about 40 to 70%. Second, they have a substantial absence of: (a) secondary structures, e.g. regions of self-complementarity (e.g. hairpins), structures formed by intramolecular hybridization events; (b) long homopolymeric stretches, e.g. polyA stretches, such that in any give homopolymeric stretch, the number of contiguous identical nucleotide bases does not exceed 5; (c) long stretches characterized by or enriched by the presence of repeating motifs, e.g., GAGAGAGA, GAAGAGAA, etc.; (d) long stretches of homopurine or homopyrimidine rich motifs; and the like.

The long oligonucleotide probes of the subject invention may be made up solely of the target specific sequence as described above, e.g., sequence designed or present which is intended for hybridization to the probe's corresponding target, or may be modified to include one or more non-target complementary domains or regions, e.g., at one or both termini of the probe, where these domains may be present to serve a number of functions, including attachment to the substrate surface, to introduce a desired conformational structure into the probe sequence, etc. One optional domain or region that may be present at one or more both termini of the long oligonucleotide probes of the subject arrays is a region enriched for the presence of thymidine bases, e.g., an oligo dT region, where the number of nucleotides in this region is typically at least 3, usually at least 5 and more usually at least 10, where the number of nucleotides in this region may be higher, but generally does not exceed about 25 and usually does not exceed about 20, where at least a substantial proportion of, if not all of, the nucleotides in this region include a thymidine base, where by substantial proportion is meant at least about 50, usually at least about 70 and more usually at least about 90 number % of all nucleotides in the oligo dT region. Certain probes of this embodiment of the subject invention, i.e., those in which the T enriched domain is an oligo dT domain, may be described by the following formula:

30 $T_n - N_m - T_k$

T is dTMP:

-7-

n and k are independently from 0 to 15, where when present n and/or k are preferably 5 to 10.

In yet other embodiments and often in addition to the above described T enriched domains, the subject probes may also include domains that impart a desired constrained structure to the probe, e.g. impart to the probe a structure which is fixed or has a restricted conformation. In many embodiments, the probes include domains which flank either end of the target specific domain and are capable of imparting a hairpin loop structure to the probe, whereby the target specific sequence is held in confined or limited conformation which enhances its binding properties with respect to its corresponding target during use. In these embodiments, the probe may be described by the following formula:



wherein:

- T is dTMP;
- N is dTMP, dGMP, dCMP or dAMP;
- m is an integer from 50 to 100;
- n and k are independently from 0 to 15, where when present n and/or k are preferably 5 to 10, where in many embodiments $k=n=5$ to 10, more preferably 10; and
- p and o are independently 5 to 20, usually 5 to 15, and more usually about 10, wherein in many embodiments $p=o=5$ to 15 and preferably 10;
- such that N_m is the target specific sequence; and
- N_o and N_p are self complementary sequences, e.g. they are complementary to each other, such that under hybridizing conditions the probe forms a hairpin loop structure in which the stem is made up of the N_o and N_p sequences and the loop is made up of the target specific sequence, i.e. N_m .

The oligonucleotide probe compositions that make up each oligonucleotide probe spot on the array will be substantially, usually completely, free of non-nucleic acids, i.e. the probe compositions will not include or be made up of non-nucleic acid biomolecules found in cells, such as proteins, lipids, and polysaccharides. In other words, the oligonucleotide spots of the arrays are substantially, if not entirely, free of non-nucleic acid cellular constituents.

The oligonucleotide probes may be nucleic acid, e.g. RNA, DNA, or nucleic acid mimetics, e.g. nucleic acids that differ from naturally occurring nucleic acids in some manner, e.g. through modified backbones, sugar residues, bases, etc., such as nucleic acids comprising non-naturally occurring heterocyclic nitrogenous bases, peptide-nucleic acids, locked nucleic acids (see Singh & Wengel, Chem. Commun. (1998) 1247-1248); and the like. In many embodiments, however, the

nucleic acids are not modified with a functionality which is necessary for attachment to the substrate surface of the array, e.g. an amino functionality, biotin, etc.

The oligonucleotide probe spots made up of the long oligonucleotides described above and present on the array may be any convenient shape, but will typically be circular, ellipsoid, oval or some other analogously curved shape. The total amount or mass of oligonucleotides present in each spot will be sufficient to provide for adequate hybridization and detection of target nucleic acid during the assay in which the array is employed. Generally, the total mass of oligonucleotides in each spot will be at least about 0.1 ng, usually at least about 0.5 ng and more usually at least about 1 ng, where the total mass may be as high as 100 ng or higher, but will usually not exceed about 20 ng and more usually will not exceed about 10 ng. The copy number of all of the oligonucleotides in a spot will be sufficient to provide enough hybridization sites for target molecule to yield a detectable signal, and will generally range from about 0.001 fmol to 10 fmol, usually from about 0.005 fmol to 5 fmol and more usually from about 0.01 fmol to 1 fmol. Where the spot is made up of two or more distinct oligonucleotides of differing sequence, the molar ratio or copy number ratio of different oligonucleotides within each spot may be about equal or may be different, wherein when the ratio of unique oligonucleotides within each spot differs, the magnitude of the difference will usually be at least 2 to 5 fold but will generally not exceed about 10 fold. Where the spot has an overall circular dimension, the diameter of the spot will generally range from about 10 to 5,000 μm , usually from about 20 to 1,000 μm and more usually from about 50 to 500 μm . The surface area of each spot is at least about 100 μm^2 , usually at least about 200 μm^2 and more usually at least about 400 μm^2 , and may be as great as 25 mm^2 or greater, but will generally not exceed about 5 mm^2 , and usually will not exceed about 1 mm^2 .

Array Features

The arrays of the subject invention are characterized by having a plurality of probe spots as described above stably associated with the surface of a solid support. The density of probe spots on the array, as well as the overall density of probe and non-probe nucleic acid spots (where the latter are described in greater detail infra) may vary greatly. As used herein, the term nucleic acid spot refers to any spot on the array surface that is made up of nucleic acids, and as such includes both probe nucleic acid spots and non-probe nucleic acid spots. The density of the nucleic acid spots on the solid surface is at least about 5/ cm^2 and usually at least about 10/ cm^2 and may be as high as 1000/ cm^2 or higher, but in many embodiments does not exceed about 1000/ cm^2 , and in these embodiments usually does not exceed about 500/ cm^2 or 400/ cm^2 , and in certain embodiments does not exceed about 300/ cm^2 . The spots may be arranged in a spatially defined and physically

addressable manner, in any convenient pattern across or over the surface of the array, such as in rows and columns so as to form a grid, in a circular pattern, and the like, where generally the pattern of spots will be present in the form of a grid across the surface of the solid support.

In the subject arrays, the spots of the pattern are stably associated with the surface of a solid support, where the support may be a flexible or rigid support. By "stably associated" it is meant that the oligonucleotides of the spots maintain their position relative to the solid support under hybridization and washing conditions. As such, the oligonucleotide members which make up the spots can be non-covalently or covalently stably associated with the support surface based on technologies well known to those of skill in the art. Examples of non-covalent association include non-specific adsorption, binding based on electrostatic (e.g., ion, ion pair interactions), hydrophobic interactions, hydrogen bonding interactions, specific binding through a specific binding pair member covalently attached to the support surface, and the like. Examples of covalent binding include covalent bonds formed between the spot oligonucleotides and a functional group present on the surface of the rigid support, e.g., -OH, where the functional group may be naturally occurring or present as a member of an introduced linking group. In many preferred embodiments, the nucleic acids making up the spots on the array surface, or at least the long oligonucleotides of the probe spots, are covalently bound to the support surface, e.g., through covalent linkages formed between moieties present on the probes (e.g., thymidine bases) and the substrate surface, etc.

As mentioned above, the array is present on either a flexible or rigid substrate. By flexible is meant that the support is capable of being bent, folded or similarly manipulated without breakage. Examples of solid materials which are flexible solid supports with respect to the present invention include membranes, flexible plastic films, and the like. By rigid is meant that the support is solid and does not readily bend, i.e., the support is not flexible. As such, the rigid substrates of the subject arrays are sufficient to provide physical support and structure to the polymeric targets present thereon under the assay conditions in which the array is employed, particularly under high throughput handling conditions. Furthermore, when the rigid supports of the subject invention are bent, they are prone to breakage.

The solid supports upon which the subject patterns of spots are presented in the subject arrays may take a variety of configurations ranging from simple to complex, depending on the intended use of the array. Thus, the substrate could have an overall slide or plate configuration, such as a rectangular or disc configuration. In many embodiments, the substrate will have a rectangular cross-sectional shape, having a length of from about 10 mm to 200 mm, usually from about 40 to 150 mm and more usually from about 75 to 125 mm and a width of from about 10 mm to 200 mm, usually from about 20 mm to 120 mm and more usually from about 25 to 80 mm, and a

thickness of from about 0.01 mm to 5.0 mm, usually from about 0.1 mm to 2 mm and more usually from about 0.2 to 1 mm. Thus, in one representative embodiment the support may have a microtitre plate format, having dimensions of approximately 12×85 mm. In another representative embodiment, the support may be a standard microscope slide with dimensions of from about 25 ×
5 75 mm.

The substrates of the subject arrays may be fabricated from a variety of materials. The materials from which the substrate is fabricated should ideally exhibit a low level of non-specific binding during hybridization events. In many situations, it will also be preferable to employ a material that is transparent to visible and/or UV light. For flexible substrates, materials of interest
10 include: nylon, both modified and unmodified, nitrocellulose, polypropylene, and the like, where a nylon membrane, as well as derivatives thereof, is of particular interest in this embodiment. For rigid substrates, specific materials of interest include: glass; plastics, e.g. polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and the like; metals, e.g. gold, platinum, and the like; etc. Also of interest are composite materials, such as glass or plastic coated
15 with a membrane, e.g. nylon or nitrocellulose, etc.

The substrates of the subject arrays comprise at least one surface on which the pattern of spots is present, where the surface may be smooth or substantially planar, or have irregularities, such as depressions or elevations. The surface on which the pattern of spots is present may be modified with one or more different layers of compounds that serve to modify the properties of the
20 surface in a desirable manner. Such modification layers, when present, will generally range in thickness from a monomolecular thickness to about 1 mm, usually from a monomolecular thickness to about 0.1 mm and more usually from a monomolecular thickness to about 0.001 mm. Modification layers of interest include: inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules and the like. Polymeric layers of interest include layers of:
25 peptides, proteins, polynucleic acids or mimetics thereof, e.g. peptide nucleic acids and the like; polysaccharides, phospholipids, polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneamines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, polyacrylamides, and the like, where the polymers may be hetero- or homopolymeric, and may or may not have separate functional moieties attached thereto, e.g. conjugated.

30 The total number of spots on the substrate will vary depending on the number of different oligonucleotide probe spots (oligonucleotide probe compositions) one wishes to display on the surface, as well as the number of non probe spots, e.g. control spots, orientation spots, calibrating spots and the like, as may be desired depending on the particular application in which the subject arrays are to be employed. Generally, the pattern present on the surface of the array will comprise

at least about 10 distinct nucleic acid spots, usually at least about 20 nucleic acid spots, and more usually at least about 50 nucleic acid spots, where the number of nucleic acid spots may be as high as 10,000 or higher, but will usually not exceed about 5,000 nucleic acid spots, and more usually will not exceed about 3,000 nucleic acid spots and in many instances will not exceed about 2,000 nucleic acid spots. In certain embodiments, it is preferable to have each distinct probe spot or probe composition be presented in duplicate, i.e., so that there are two duplicate probe spots displayed on the array for a given target. In certain embodiments, each target represented on the array surface is only represented by a single type of oligonucleotide probe. In other words, all of the oligonucleotide probes on the array for a give target represented thereon have the same sequence. In certain embodiments, the number of spots will range from about 200 to 1200. The number of probe spots present in the array will typically make up a substantial proportion of the total number of nucleic acid spots on the array, where in many embodiments the number of probe spots is at least about 50 number %, usually at least about 80 number % and more usually at least about 90 number % of the total number of nucleic acid spots on the array. As such, in many embodiments the total number of probe spots on the array ranges from about 50 to 20,000, usually from about 100 to 10,000 and more usually from about 200 to 5,000.

In the arrays of the subject invention (particularly those designed for use in high throughput applications, such as high throughput analysis applications), a single pattern of oligonucleotide spots may be present on the array or the array may comprise a plurality of different oligonucleotide spot patterns, each pattern being as defined above. When a plurality of different oligonucleotide spot patterns are present, the patterns may be identical to each other, such that the array comprises two or more identical oligonucleotide spot patterns on its surface, or the oligonucleotide spot patterns may be different, e.g. in arrays that have two or more different types of target nucleic acids represented on their surface, e.g., an array that has a pattern of spots corresponding to human genes and a pattern of spots corresponding to mouse genes. Where a plurality of spot patterns are present on the array, the number of different spot patterns is at least 2, usually at least 6, more usually at least 24 or 96, where the number of different patterns will generally not exceed about 384.

Where the array comprises a plurality of oligonucleotide spot patterns on its surface, preferably the array comprises a plurality of reaction chambers, wherein each chamber has a bottom surface having associated therewith an pattern of oligonucleotide spots and at least one wall, usually a plurality of walls surrounding the bottom surface. See e.g. U.S. Patent No. 5,545,531, the disclosure of which is herein incorporated by reference. Of particular interest in many embodiments are arrays in which the same pattern of spots is reproduced in 24 or 96 different reaction chambers across the surface of the array.

Within any given pattern of spots on the array, there may be a single spot that corresponds to a given target or a number of different spots that correspond to the same target, where when a plurality of different spots are present that correspond to the same target, the probe compositions of each spot that corresponds to the same target may be identical or different. In other words, a plurality of different targets are represented in the pattern of spots, where each target may correspond to a single spot or a plurality of spots, where the oligonucleotide probe composition among the plurality of spots corresponding to the same target may be the same or different. Where a plurality of spots (of the same or different composition) corresponding to the same target is present on the array, the number of spots in this plurality will be at least about 2 and may be as high as 10, but will usually not exceed about 5. As mentioned above, however, in many preferred embodiments, however, any given target nucleic acid is represented by only a single type of probe spot, which may be present only once or multiple times on the array surface, e.g. in duplicate, triplicate etc.

The number of different targets represented on the array is at least about 2, usually at least about 10 and more usually at least about 20, where in many embodiments the number of different targets, e.g. genes, represented on the array is at least about 50 and more usually at least about 100. The number of different targets represented on the array may be as high as 5,000 or higher, but in many embodiments will usually not exceed about 3,000 and more usually will not exceed about 2,500. A target is considered to be represented on an array if it is able to hybridize to one or more probe compositions on the array.

Another feature of the present invention is that the relative binding efficiencies of each of the distinct long oligonucleotide probes for their respective targets is substantially the same, such that the binding efficiency of any two different long oligonucleotide probes on the arrays for their respective targets does not vary by more than about 20 fold, usually by not more than about 15 fold and more usually by not more than about 10 fold, where in many embodiments the binding efficiencies do not vary by more than about 5 fold and preferably by not more than about 3 fold.

In certain preferred embodiments of the invention, each of the probe spots in the array comprising the long oligonucleotide probe compositions correspond to the same kind of gene; i.e., genes that all share some common characteristic or can be grouped together based on some common feature, such as species of origin, tissue or cell of origin, functional role, disease association, etc. In this embodiment, each of the different target nucleic acids that corresponds to the different probe spots on the array are of the same type, i.e., that are coding sequences of the same type of gene. As such, the arrays of this embodiment of the subject invention will be of a specific array type. A variety of specific array types are provided by the subject invention. Specific

array types of interest include: human, cancer, apoptosis, cardiovascular, cell cycle, hematology, mouse, human stress, mouse stress, oncogene and tumor suppressor, cell-cell interaction, cytokine and cytokine receptor, rat, rat stress, blood, mouse stress, neurobiology, and the like.

5 With respect to the oligonucleotide probes that correspond to a particular type or kind of gene, type or kind can refer to a plurality of different characterizing features, where such features include: species specific genes, where specific species of interest include eukaryotic species, such as mice, rats, rabbits, pigs, primates, humans, etc.; function specific genes, where such genes include oncogenes, apoptosis genes, cytokines, receptors, protein kinases, etc.; genes specific for or involved in a particular biological process, such as apoptosis, differentiation, stress response, aging, proliferation, etc.; cellular mechanism genes, e.g. cell-cycle, signal transduction, metabolism of toxic compounds, etc.; disease associated genes, e.g. genes involved in cancer, schizophrenia, diabetes, high blood pressure, atherosclerosis, viral-host interaction and infection diseases, etc.; location specific genes, where locations include organ, such as heart, liver, prostate, lung etc., tissue, such as nerve, muscle, connective, etc., cellular, such as axonal, lymphocytic, etc, or
10 subcellular locations, e.g. nucleus, endoplasmic reticulum, Golgi complex, endosome, lysosome, peroxisome, mitochondria, cytoplasm, cytoskeleton, plasma membrane, extracellular space, chromosome-specific genes; specific genes that change expression level over time, e.g. genes that are expressed at different levels during the progression of a disease condition, such as prostate genes which are induced or repressed during the progression of prostate cancer.

20 In addition to the oligonucleotide spots comprising the oligonucleotide probe compositions (i.e. oligonucleotide probe spots), the subject arrays may comprise one or more additional spots of polynucleotides or nucleic acid spots which do not correspond to target nucleic acids as defined above, such as target nucleic acids of the type or kind of gene represented on the array in those embodiments in which the array is of a specific type. In other words, the array may comprise one or
25 more non probe nucleic acid spots that are made of non "unique" oligonucleotides or polynucleotides, i.e., common oligonucleotides or polynucleotides. For example, spots comprising genomic DNA may be provided in the array, where such spots may serve as orientation marks. Spots comprising plasmid and bacteriophage genes, genes from the same or another species which are not expressed and do not cross hybridize with the cDNA target, and the like, may be present
30 and serve as negative controls. In addition, spots comprising a plurality of oligonucleotides complimentary to housekeeping genes and other control genes from the same or another species may be present, which spots serve in the normalization of mRNA abundance and standardization of hybridization signal intensity in the sample assayed with the array. Orientation spots may also be included on the array, where such spots serve to simplify image analysis of hybrid patterns. Other

types of spots include spots for calibration or quantitative standards, controls for integrity of RNA template (targets), controls for efficiency steps in target preparation (such as efficiency of labeling, purification and hybridization), etc. These latter types of spots are distinguished from the oligonucleotide probe spots, i.e. they are non-probe spots.

5

Array Preparation

The subject arrays can be prepared using any convenient means. One means of preparing the subject arrays is to first synthesize the oligonucleotides for each spot and then deposit the oligonucleotides as a spot on the support surface. The oligonucleotides may be prepared using any
10 convenient methodology, where chemical synthesis procedures using phosphoramidite or analogous protocols in which individual bases are added sequentially without the use of a polymerase, e.g. such as is found in automated solid phase synthesis protocols, and the like, are of particular interest, where such techniques are well known to those of skill in the art.

In determining the specific oligonucleotides of the probe compositions, the oligonucleotide
15 should be chosen so that is capable of hybridizing to a region of the target nucleic acid or gene having a sequence unique to that gene. Different methods may be employed to choose the specific region of the gene to which the oligonucleotide probe is to hybridize. Thus, one can use a random approach based on availability of a gene of interest. However, instead of using a random approach which is based on availability of a gene of interest, a rational design approach may also be
20 employed to choose the optimal sequence for the hybridization array. Preferably, the region of the gene that is selected in preparing the oligonucleotide probe is chosen based on the following criteria. First, the sequence that is chosen as the target specific sequence should yield an oligonucleotide probe that does not cross-hybridize with, or is homologous to, any other oligonucleotide probe for other spots present on the array that do not correspond to the target gene.
25 Second, the sequence should be chosen such that the oligonucleotide probe has a low homology to a nucleotide sequence found in any other gene, whether or not the gene is to be represented on the array from the same species of origin. As such, sequences that are avoided include those found in: highly expressed gene products, structural RNAs, repeated sequences found in the RNA sample to be tested with the array and sequences found in vectors. A further consideration is to select
30 sequences which provide for minimal or no secondary structure, structure which allows for optimal hybridization but low non-specific binding, equal or similar thermal stabilities, and optimal hybridization characteristics. A final consideration is to select probe sequences that give rise to probes which efficiently hybridize to their corresponding target and do not suffer from substantial non-specific hybridization events. Finally, all of the probe sequences on the array are preferably

chosen such that they exhibit substantially the same hybridization efficiency to their corresponding probes, where the difference in hybridization efficiency between any two probes and their corresponding targets preferably does not exceed about 10 fold, more preferably does not exceed about 5 fold and most preferably does not exceed about 3 fold.

5 Probes meeting the above criteria can be designed or identified using any convenient protocol. A representative protocol includes the following algorithm which is part of the present invention. In selecting probes according to this representative algorithm or process, a unique gene-specific or target specific sequence (one or more regions per gene) is first identified based on a sequence homology search algorithm described in detail in copending application serial no.
10 09/053,375, the disclosure of which is herein incorporated by reference. In this step, the sequence of all genes represented on the to be produced array and all sequences deposited in GenBank are searched in order to select mRNA fragments which are unique for each mRNA or target to be represented on the array. A unique sequence is defined as a sequence which at least does not have significant homology to any other sequence on the array. For example, where one is interested in
15 identifying suitable 80 base long unique probes, sequences which do not have homology of more than about 80% to any consecutive 40 base segment of any of the other probes on the array are selected. This step typically results in a reduced population of candidate probe sequences as compared to the initial population of possible sequences identified for each specific target.

Of this reduced population of candidate sequences, screening criteria are employed to
20 exclude non-optimal sequences, where sequences that are excluded or screened out in this step include: (a) those with strong secondary structure or self-complementarity (for example long hairpins); (b) those with very high (more than 70%) or very low (less than 40%) GC content; (c) those with long stretches (more than 6) of identical consecutive bases or long stretches of sequences enriched in some motifs, purine or pyrimidine stretches or particular bases, like GAGAGAGA...,
25 GAAGAGAA; and the like. This step results in a further reduction in the population of candidate probe sequences.

In the next step, sequences are selected that have similar melting temperatures or thermodynamic stability which will provide similar performance in hybridization assays with target nucleic acids. Of interest is the identification of probes that can participate in duplexes whose
30 melting temperature exceeds 65, usually at least about 75 and more usually at least about 80°C.

The final step in this representative design process is to select from the remaining sequences those sequences which provide for low levels of non-specific hybridization and similar high efficiency hybridization with complementary target molecules. This final selection is accomplished by practicing the following steps:

1. The remaining set of probes which is identified for each target using the above steps, where this remaining set typically includes at least 1 potential probe, usually at least 2 potential probes and more usually at least 3 potential probes, are experimentally characterized for their hybridization efficiency and propensity to participate in non-specific hybridization events using the following protocol.
2. First, an array of at least a portion of the candidate probes for each target to be represented on the final array is produced. For example, where three candidate probes have been identified for a particular target sequence, these probes are attached to the surface of a solid support, along with candidate probes for other targets, to produce a test probe array.
3. Next, a normalization control target set is prepared, wherein each target in the set is complementary to one probe sequence in the array and the various target constituents of the set are mixed in similar or identical amounts. The number of targets in the set of control targets is usually less than the set of probes in the array. Usually the number of targets in the control set is between 50% and 90%, but can be between 10 and 100%, of the number of test probes on the array surface. As such, not all of the probe sequences on the test array will have a corresponding or complementary target in the target control set. For example, where three different candidate probes have been identified for each of 10 different mRNA targets, a test probe array of 30 different oligonucleotide probes is prepared. Next, a control set of target nucleic acids which includes targets that correspond to 5 of the 10 different mRNA targets represented on the array is produced, where the control set includes a target that is complementary to each different probe corresponding to 1 of the 5 different mRNAs represented in the control target set, i.e. the control target set includes 15 different targets--1 target for each of the 15 probes on the array that correspond to the 5 different mRNAs represented in the control target set. (While the above procedure has been described in terms of using a target population that corresponds to less than all of the probes on the array so that non-specific hybridization can be determined, other protocols also may be employed. For example, one may use a population of targets that corresponds to all of the probes on the array, where at least a portion of the targets are distinguishable from the remaining portion or portions, e.g. by label, mass etc. Following hybridization, the targets hybridized to each probe can be detected and both the efficiency of the probe for its true target and its propensity for non-specific hybridization can be determined).
4. Following generation of the control set of targets, the control set is hybridized with the test probe array under stringent conditions and hybridization signals are detected. The intensity

of the signal for those probes which have a corresponding labeled complementary target in the hybridization solution is used as a measure for determining the hybridization efficiency of that probe, as well as differences in hybridization efficiency of different candidate probes for different targets. For those probes on the array which do not have complementary
5 labeled target sequences in control set, the intensity of hybridization signal generated by each of these probes is used to identify the level of non-specific hybridization that characterizes these probes.

5. The above steps are repeated with one or more additional control sets of target nucleic acids in order to get comprehensive information concerning the hybridization efficiency and level
10 of non-specific hybridization for each candidate of the candidate probes on the array. The number of different sets of control targets that are employed in this process is generally at least two, more commonly at least four and most commonly at least ten.
6. From the above steps, probe sequences meeting the following criteria are identified for use as long oligonucleotide probes in the arrays of the subject invention. First, candidate probes
15 that exhibit a high efficiency of hybridization for their corresponding targets are identified. In many embodiments, candidate probes having substantially the same hybridization efficiency for the respective targets are identified, where any two probes to different targets have substantially the same hybridization efficiency for their respective targets if the differences in hybridization efficiency of the two probes does not exceed 10-fold, where
20 differences of less than about 5-fold and often less than about 3-fold are preferred. Of these identified probes, probes that show substantial cross hybridization or non-specific hybridization are excluded, where a probe that shows non-specific hybridization of up to at least 5-fold, more commonly 20-fold and most commonly 50-fold less than the level of gene-specific hybridization between the probe and its corresponding target are excluded in
25 this step. In other words, in the above assay hybridizations, those probes that exhibit a signal that is at within 5-fold less, usually at least 20-fold less and more usually within 50-fold less of the signal generated by probes and their complementary targets are excluded as being probes with unacceptably high propensities for participating in non-specific hybridization events.

30

The above algorithm or process is used to design the long oligonucleotide probes that are present on the arrays of the subject invention. Steps 1 to 6 can be repeated if, in the first round of selection for particular targets no array candidate probes were identified. Once the design or

sequence of the probes is identified, the long oligonucleotide probes may be synthesized according to any convenient protocol, as mentioned above, e.g. via phosphoramidite processes.

Following synthesis of the subject long oligonucleotide probes, the probes are stably associated with the surface of the solid support. This portion of the preparation process typically involves deposition the probes, e.g. a solution of the probes, onto the surface of the substrate, where the deposition process may or may not be coupled with a covalent attachment step, depending on how the probes are to be stably attached to the substrate surface, e.g. via electrostatic interactions, covalent bonds, etc. The prepared oligonucleotides may be spotted on the support using any convenient methodology, including manual techniques, e.g. by micro pipette, ink jet, pins, etc., and automated protocols. Of particular interest is the use of an automated spotting device, such as the BioGrid Arrayer (Biorobotics).

Where desired, the long oligonucleotides can be covalently bonded to the substrate surface using a number of different protocols. For example, functionally active groups such as amino, etc., can be introduced onto the 5' or 3' ends of the oligonucleotides, where the introduced functionalities are then reacted with active surface groups on the substrate to provide the covalent linkage. In certain preferred embodiments, the long oligonucleotide probes are covalently bonded to the surface of the substrate using the following protocol. In this process, the probes are covalently attached to the substrate surface under denaturing conditions. Typically, a denaturing composition of each probe is prepared and then deposited on the substrate surface. By denaturing composition is meant that the probe molecules present in the composition are not participating in secondary structures, e.g. through self-hybridization or hybridization to other molecules in the composition. The denaturing composition, typically a fluid composition, may be any composition which inhibits the formation of hydrogen bonds between complementary nucleotide bases. Thus, compositions of interest are those that include a denaturing agent, e.g. urea, formamide, sodium thiocyanate, etc., as well as solutions having a high pH, e.g. 12 to 13.5, usually 12.5 to 13, or a low pH, e.g. 1 to 4, usually 1 to 3; and the like. In many preferred embodiments, the composition is a strongly alkaline solution of the long oligonucleotide, where the composition comprises a base, e.g. sodium hydroxide, lithium hydroxide, potassium hydroxide, ammonium hydroxide, tetramethyl ammonium hydroxide, ammonium hydroxide, etc, in sufficient amounts to impart the desired high pH to the composition, e.g. 12.5 to 13.0. The concentration of long oligonucleotide in the composition typically ranges from about 0.1 to 10 μ M, usually from about 0.5 to 5 μ M. Following deposition of the denaturing composition of the long oligonucleotide probe onto the substrate surface, the deposited probe is exposed to UV radiation of sufficient wavelength, e.g. from 250 to 350 nm, to cross link the deposited probe to the surface of the substrate. The irradiation wavelength for this

process typically ranges from about 50 to 1000 mJoules, usually from about 100 to 500 mJoules, where the duration of the exposure typically lasts from about 20 to 600 sec, usually from about 30 to 120 sec.

5 The above protocol for covalent attachment results in the random covalent binding of the long oligonucleotide probe to the substrate surface by one or more attachment sites on the probe, where such attachment may optionally be enhanced through inclusion of oligodT regions at one or more ends of the oligonucleotides, as discussed supra. An important feature of the above process is that reactive moieties, e.g. amino, that are not present on naturally occurring oligonucleotides are not employed in the subject methods. As such, the subject methods are suitable for use with
10 oligonucleotides that do not include moieties that are not present on naturally occurring nucleic acids.

The above described covalent attachment protocol may be used with a variety of different types of substrates. Thus, the above described protocols can be employed with solid supports, such as glass, plastics, membranes, e.g. nylon, and the like. The surfaces may or may not be modified.
15 For example, the nylon surface may be charge neutral or positively charged, where such substrates are available from a number of commercial sources. For glass surfaces, in many embodiments the glass surface is modified, e.g. to display reactive functionalities, such as amino, phenyl isothiocyanate, etc.

20 METHODS OF USING THE SUBJECT ARRAYS

The subject arrays find use in a variety of different applications in which one is interested in detecting the occurrence of one or more binding events between target nucleic acids and probes on the array and then relating the occurrence of the binding event(s) to the presence of a target(s) in a sample. In general, the device will be contacted with the sample suspected of containing the target
25 under conditions sufficient for binding of any target present in the sample to complementary oligonucleotides present on the array. Generally, the sample will be a fluid sample and contact will be achieved by introduction of an appropriate volume of the fluid sample onto the array surface, where introduction can be through delivery ports, direct contact, deposition, and the like.

30 *Generation of Labeled Target*

Targets may be generated by methods known in the art. mRNA can be labeled and used directly as a target, or converted to a labeled cDNA target. Alternatively, an excess of synthetic labeled oligonucleotide target which is complementary to the probes on the array can be hybridized with the mRNA, followed by separation of any unbound target from the hybridized fraction or

isolation of the hybridized fraction. The hybridized fraction can then hybridized to the array to reveal the expression pattern of the cellular source from which the mRNA was derived. Usually, mRNA is labeled non-specifically (randomly) directly using chemically, photochemically or enzymatically activated labeling compounds, such as photobiotin (Clontech, Palo Alto, CA), Dig-
5 Chem-Link (Boehringer), and the like. In another way, mRNA target can be labeled specifically in the sequences which are complementary to the probes. This specific labeling can be achieved by using covalent or non-covalent binding of additional labeled oligonucleotides (or mimetics) to the target sequences which flank the probe complementary sequence or the complementary probe sequence. The hybridized fraction of labeled oligonucleotides with mRNA can be purified or
10 separated from the non-hybridized fraction and then hybridized to the array. Generally, methods for generating labeled cDNA probes include the use of oligonucleotide primers. Primers that may be employed include oligo dT, random primers, e.g. random hexamers and gene specific primers, as described in PCT/US98/10561, the disclosure of which is herein incorporated by reference.

Where gene specific primers are employed, the gene specific primers are preferably those
15 primers that correspond to the different oligonucleotide spots on the array. Thus, one will preferably employ gene specific primers for each different oligonucleotide that is present on the array, so that if the gene is expressed in the particular cell or tissue being analyzed, labeled target will be generated from the sample for that gene. In this manner, if a particular gene present on the array is expressed in a particular sample, the appropriate target will be generated and subsequently
20 identified. For each target represented on the array, a single gene specific primer may be employed or a plurality of different gene specific primers may be employed, where when a plurality are used to produce the target, the number will generally not exceed about 3. Generally, in preparing the target from template nucleic acid, e.g. mRNA, the gene specific primers will hybridize to a region of the template that is downstream from the region to which the probes are homologous, e.g. to
25 which the probes are complementary or have the same sequence. The distance from oligonucleotide probe sequence and primer binding site generally does not exceed about 500 nt, usually does not exceed about 300 nt and more usually does not exceed about 200 nt. However, in certain embodiments the gene specific primers may be partially or completely complementary to the oligonucleotide probes. The cDNA probe can be further amplified by PCR or can be converted
30 (linearly amplified) using phage coded RNA polymerase transcription of dsDNA. See PCT/US98/1056, the disclosure of which is herein incorporated by reference.

In many embodiments, the target that is generated in this step is a linear target which is devoid of any secondary structure, e.g. as produced by target intramolecular interactions such as hydrogen bonds. However, in certain embodiments, it may be desirable to generate a

conformationally restricted to constrained target, e.g. a target that forms a hairpin loop structure under the hybridization conditions in which the target is employed. One means of producing hairpin loop targets is to employ primers that include an anchoring sequence in addition to priming sequence in the enzymatic target generation step. The anchoring domain of the primer, which is 5' of the priming domain, is a domain that is complementary to a region of the first strand cDNA distal to the 5' end that is generated during target synthesis, where the 5' distal region to which the anchor is complementary is sufficiently separated from the 5' end of the cDNA such that the cDNA forms a hairpin loop structure in which the anchor sequence of the 5' distal region to which the anchor sequence is complementary form the stem structure. The sequence of the anchor domain of the primer is typically chosen to provide for a loop that ranges in size from about 20 to 200 nt, usually from about 30 to 100 nt and more usually from about 40 to 80 nt. The primers used to generate these hairpin loop targets are described by the following formula:



wherein

- 15 N is dGMP, dCMP, dAMP and dTMP;
p is an integer ranging from 12 to 35, usually from 15 to 30 and more usually from 18 to 25, such that N_p is the priming domain of the primer, and may be a gene specific domain, as described above, or an oligo dT domain; and
x is an integer ranging from 3 to 30, usually from 5 to 20 and more usually from 5 to 15,
20 wherein N_x is the anchor domain and is complementary to a 5' distal portion of the first strand cDNA that is complementary to the mRNA of interest which is to be represented as target.

- A variety of different protocols may be used to generate the labeled target nucleic acids, as is known in the art, where such methods typically rely in the enzymatic generation of the labeled target using the initial primer. Labeled primers can be employed to generate the labeled target.
25 Alternatively, label can be incorporated during first strand synthesis or subsequent synthesis labeling or amplification steps, including chemical or enzymatic labeling steps, in order to produce labeled target. Representative methods of producing labeled target are disclosed in PCT/US98/10561, the disclosure of which is herein incorporated by reference.

30 *Hybridization and Detection*

As mentioned above, following preparation of the target nucleic acid from the tissue or cell of interest, the target nucleic acid is then contacted with the array under hybridization conditions, where such conditions can be adjusted, as desired, to provide for an optimum level of specificity in view of the particular assay being performed. Suitable hybridization conditions are well known to

those of skill in the art and reviewed in Maniatis et al, *supra* and WO 95/21944. Of particular interest in many embodiments is the use of stringent conditions during hybridization, i.e. conditions that are optimal in terms of rate, yield and stability for specific probe-target hybridization and provide for a minimum of non-specific probe/target interaction. Stringent conditions are known to those of skill in the art. In the present invention, stringent conditions are typically characterized by temperatures ranging from 15 to 35, usually 20 to 30 °C less than the melting temperature of the probe target duplexes, which melting temperature is dependent on a number of parameters, e.g. temperature, buffer compositions, size of probes and targets, concentration of probes and targets, etc. As such, the temperature of hybridization typically ranges from about 55 to 70, usually from about 60 to 68 °C. In the presence of denaturing agents, the temperature may range from about 35 to 45, usually from about 37 to 42 °C. The stringent hybridization conditions are further typically characterized by the presence of a hybridization buffer, where the buffer is characterized by one or more of the following characteristics: (a) having a high salt concentration, e.g. 3 to 6 × SSC (or other salts with similar concentrations); (b) the presence of detergents, like SDS (from 0.1 to 20%), triton X100 (from 0.01 to 1%), monidet NP40 (from 0.1 to 5%) etc.; (c) other additives, like EDTA (typically from 0.1 to 1 μM), tetramethylammonium chloride; (d) accelerating agents, e.g. PEG, dextran sulfate (5 to 10 %), CTAB, SDS and the like; (e) denaturing agents, e.g. formamide, urea etc.; and the like.

In analyzing the differences in the population of labeled target nucleic acids generated from two or more physiological sources using the arrays described above, in certain embodiments each population of labeled target nucleic acids are separately contacted to identical probe arrays or together to the same array under conditions of hybridization, preferably under stringent hybridization conditions, such that labeled target nucleic acids hybridize to complementary probes on the substrate surface. In yet other embodiments, labeled target nucleic acids are combined with a distinguishably labeled standard or control target nucleic acids followed by hybridization of the combined populations to the array surface, as described in application serial no. 09/298,361; the disclosure of which is herein incorporated by reference.

Where all of the target sequences comprise the same label, different arrays will be employed for each physiological source (where different could include using the same array at different times). Alternatively, where the labels of the targets are different and distinguishable for each of the different physiological sources being assayed, the opportunity arises to use the same array at the same time for each of the different target populations. Examples of distinguishable labels are well known in the art and include: two or more different emission wavelength fluorescent dyes, like Cy3 and Cy5, two or more isotopes with different energy of emission, like ³²P and ³³P, gold or silver

particles with different scattering spectra, labels which generate signals under different treatment conditions, like temperature, pH, treatment by additional chemical agents, etc., or generate signals at different time points after treatment. Using one or more enzymes for signal generation allows for the use of an even greater variety of distinguishable labels, based on different substrate specificity of enzymes (alkaline phosphatase/peroxidase).

Following hybridization, non-hybridized labeled nucleic acid is removed from the support surface, conveniently by washing, generating a pattern of hybridized nucleic acid on the substrate surface. A variety of wash solutions are known to those of skill in the art and may be used.

The resultant hybridization patterns of labeled nucleic acids may be visualized or detected in a variety of ways, with the particular manner of detection being chosen based on the particular label of the target nucleic acid, where representative detection means include scintillation counting, autoradiography, fluorescence measurement, colorimetric measurement, light emission measurement, light scattering, and the like.

Following detection or visualization, the hybridization patterns may be compared to identify differences between the patterns. Where arrays in which each of the different probes corresponds to a known gene are employed, any discrepancies can be related to a differential expression of a particular gene in the physiological sources being compared.

The provision of appropriate controls on the arrays permits a more detailed analysis that controls for variations in hybridization conditions, cross-hybridization, non-specific binding and the like. Thus, for example, in a preferred embodiment, the hybridization array is provided with normalization controls as described supra. These normalization controls are probes complementary to control target sequences added in a known concentration to the sample. Where the overall hybridization conditions are poor, the normalization controls will show a smaller signal reflecting reduced hybridization. Conversely, where hybridization conditions are good, the normalization controls will provide a higher signal reflecting the improved hybridization. Normalization of the signal derived from other probes in the array to the normalization controls thus provides a control for variations in hybridization conditions. Normalization control is also useful to adjust (e.g. correct) for differences which arise from the array quality, the mRNA sample quality, efficiency of first-strand synthesis, etc. Typically, normalization is accomplished by dividing the measured signal from the other probes in the array by the average signal produced by the normalization controls. Normalization may also include correction for variations due to sample preparation and amplification. Such normalization may be accomplished by dividing the measured signal by the average signal from the sample preparation/ amplification control probes. The resulting values may be multiplied by a constant value to scale the results.

In certain embodiments, normalization controls are often unnecessary for useful quantification of a hybridization signal. Thus, where optimal probes have been identified, the average hybridization signal produced by the selected optimal probes provides a good quantified measure of the concentration of hybridized nucleic acid. However, normalization controls may still
5 be employed in such methods for other purposes, e.g. to account for array quality, mRNA sample quality, etc.

UTILITY

The subject methods find use in, among other applications, differential gene expression
10 assays. Thus, one may use the subject methods in the differential expression analysis of: (a) diseased and normal tissue, *e.g.* neoplastic and normal tissue, (b) different tissue or tissue types; (c) developmental stage; (d) response to external or internal stimulus; (e) response to treatment; and the like. The subject arrays therefore find use in broad scale expression screening for drug
discovery, diagnostics and research, as well as studying the effect of a particular active agent on the
15 expression pattern of genes in a particular cell, where such information can be used to reveal drug toxicity, carcinogenicity, etc., environmental monitoring, disease research and the like.

KITS

Also provided are kits for performing analyte binding assays using the subject devices,
20 where kits for carrying out differential gene expression analysis assays are preferred. Such kits according to the subject invention will at least comprise the subject arrays. The kits may further comprise one or more additional reagents employed in the various methods, such as primers for generating target nucleic acids, dNTPs and/or rNTPs, which may be either premixed or separate,
one or more uniquely labeled dNTPs and/or rNTPs, such as biotinylated or Cy3 or Cy5 tagged
25 dNTPs, gold or silver particles with different scattering spectra, or other post synthesis labeling reagent, such as chemically active derivatives of fluorescent dyes, enzymes, such as reverse transcriptases, DNA polymerases, RNA polymerases, and the like, various buffer mediums, *e.g.* hybridization and washing buffers, prefabricated probe arrays, labeled probe purification reagents and components, like spin columns, etc., signal generation and detection reagents, *e.g.* streptavidin-
30 alkaline phosphatase conjugate, chemifluorescent or chemiluminescent substrate, and the like.

The following examples are offered by way of illustration and not by way of limitation.

EXPERIMENTAL

In the following examples, all percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

5 **Example 1 - Generation of ^{32}P -labeled hybridization target.**

Step A. cDNA Synthesis/Labeling Procedure

The 10- μl reaction described below convert 1 μg of synthetic control RNA into ^{32}P -labeled first-strand cDNA.

For each labeling reaction:

- 10 1. Prepare enough master mix for all labeling reactions and 1 extra reaction to ensure sufficient volume. For each 10- μl labeling reaction, mix the following reagents:

2 μl	5 \times First-strand buffer (250 μM Tris-HCl pH8.3; 375 mM KCl; 15 mM MgCl_2)
1 μl	10 \times dNTP mix (500 μM dGTP, 500 μM dCTP, 500 μM dTTP, 5 μM dATP)
4 μl	$[\alpha\text{-}^{32}\text{P}]\text{dATP}$ (Amersham, 2500 Ci/mmol, 10 mCi/ml)
15 1 μl	MMLV reverse transcriptase (Amersham, 200 units/ μl)

8 μl Final volume

2. Combine the following in a 0.5-ml PCR test tube:

	1 μg (1 μl)	control s64 RNA
20	GGCCA GGATACAAA GCCTTACAGG ACTTCCTCCT CAGTGTGCAG ATGTGCCCAG GTAATCGAGA CACTTACTTT CACCTGCTTC AGACTCTGAA GAGGCTAGAT CGGAGGGATG AGGCCACTGC ACTCTGGTGG AGGCTGGAGG CCCAACTAA GGGGTCACAT GAAGATGCTC TGTGGTCTCT CCCCCTGTAC CTAGAAAGCT ATTTGAGCTG GATCCGTCCC TCTGATCGTG ACGCCTTCCT TGAAGAATTT CGGACATCTC TGCCAAAGTC TTGTGACCTG TAGCTGCC (SEQ ID NO: 01)	
25	1 μl	gene-specific primer s64 (0.2 μM)
	CGGCCAGGATACCAAAGCCTTACAG (SEQ ID NO: 02)	

- The control s64 RNA provided above was synthesized by T7 transcription from cDNA fragment corresponding to the human DNA repair protein XRCC9 (GB accession number U70310) as described in more details in patent application serial no. 09/298,361, the disclosure of which is
30 herein incorporated by reference.

3. Add ddH_2O to a final volume of 3 μl .
4. Mix contents and spin the tubes briefly in a microcentrifuge.
5. Incubate the tubes in preheated PCR thermocycler at 70°C for 2 min.
6. Reduce temperature in thermocycle down to 50°C and incubate for 2 min.
35 7. Add 8 μl of master mix to each reaction test tube.
8. Mix the contents of the test tubes by gentle pipetting.

9. Incubate the tubes in PCR thermocycler for 20 min at 50°C.
10. Stop the reaction by adding 1 µl of 10× termination mix (0.1 M EDTA, 1 mg/ml glycogen).

Step B. Column Chromatography

- 5 To purify the ³²P-labeled cDNAs from unincorporated ³²P-labeled nucleotides and small (<0.1- kb) cDNA fragments, follow this procedure for each test tube:
 1. Remove CHROMA SPIN-200 column (CLONTECH) from refrigerator and warm up at room temperature for about 1 hour. Invert the column several times to completely re-suspend the gel matrix.
- 10 Note: Check for air bubbles in the column matrix. If bubbles are visible, re-suspend the matrix in the in the column buffer (ddH₂O) by inverting the column again.
 2. Remove the bottom cap from the column, and then slowly remove the top cap.
 3. Place the column into a 1.5-ml microcentrifuge tube.
 4. Let the water drain through the column by gravity flow until you can see the surface of the gel
- 15 beads in the column matrix. The top of the column matrix should be at 0.75-ml mark on the wall of the column. If the column contains less matrix, adjust the volume of the matrix to 0.75-ml mark using matrix from another column.
 5. Discard the collected water and proceed with purification.
 6. Carefully and slowly apply the sample to the center of the gel bed's flat surface and allow sample
- 20 to be fully absorbed into the resin bed before proceeding to the next step. Do not allow any sample to flow along the inner wall of the column.
 7. Apply 25 µl of ddH₂O and allow the water to completely drain out of the column.
 8. Apply 200 µl of ddH₂O and allow the buffer to completely drain out of the column until there is no liquid left above the resin bed.
- 25 9. Transfer column to a clean 1.5-ml microcentrifuge tube.
 10. To collect the first fraction add 100 µl of ddH₂O to the column and allow the water to completely drain out of the column.
 11. To collect the second, third and fourth fractions repeat steps 9-10.
 12. Place the tubes with fractions 1-4 in a scintillation counter empty vials (do not add scintillation
- 30 cocktail to the tubes or vials), and obtain Cerenkov counts for each fraction. Count the entire sample in the tritium channel.
 13. Pool the fractions (usually fractions 2-3) which show the highest Cerenkov counts. Waist column and the fractions (usually fraction 1 and 4) which show less than 10% counts from peak fractions. Total incorporation into peak fractions should be 2-5×10⁶ cpm.

Example 2. Preparation of Aminoethyl-glass.

1. Prepare wash solution: to get 2 liters, dissolve 200g NaOH in 600ml water and make up volume to 1 liter (20% w/v). To this solution add 1 liter ethanol. This makes 10% NaOH in 50% EtOH. Wash glass in this solution on orbital shaker overnight. (slides are placed in rack)
2. Transfer rack(s) with slides into bath with MilliQ water and wash on shaker for 15-20 min, repeat this step one more time.
3. Transfer slides into bath with acetone and wash on shaker for 15-20 min. Repeat this step two more times. Dispose acetone from first wash and keep acetone from 2nd and 3rd washes. (When doing this procedure again, use 2nd wash as first, 3rd as second and for the 3rd wash use fresh acetone.
4. Prepare in advance 5% solution of water in acetone (5% water – 95% acetone).
5. During last wash step prepare 0.5% solution of aminoethyltriethoxysilane (Sigma, cat No A3648) in acetone-water mixture from step 4.
6. Transfer slides from last acetone wash into silanization solution and incubate for 2 hours at room temperature on orbital shaker.
7. Transfer slides into MilliQ water and wash for 20 minutes.
8. Transfer slides into acetone and wash for 20 min, repeat this step 2 more times. These acetone washes are to be disposed.
9. Preheat oven at 110°C
10. Remove rack with slides from the last acetone wash and transfer it into preheated oven. As some acetone still remains on slides and on rack's surfaces, the smelt becomes quite intensive. Exhaust duct should be open after putting slides into oven and may be closed after first 30 minutes of baking.
11. Program oven to bake slides at 110°C for 3 hours and then shut down or cool down to room temperature. It is convenient to do this step overnight.
12. After baking is over, slides are ready for printing using "thiocyanate method". If the printing will not be done right away, slides may be kept in clean boxes inside dry cabinets.

The following steps are for preparation of PDITC-slides.

1. Prepare a mixture of Pyridine and Dimethylformamide (10% pyridine and 90% DMF). Prepare only as much as necessary. This mixture cannot be stored.
2. Dissolve 1,4-Phenylenediisothiocyanate in the Pyridine-DMF mixture at 0.1%

concentration (1g per liter) on stirrer. Prepare this solution only as much as necessary and only when ready to proceed with next steps. This solution cannot be stored. The solution should be light yellow-green in color.

3. Pour the solution in a tray and transfer tray(s) with amino-modified slides into the solution.
- 5 Close the tray with the lid and shake on orbital shaker at low speed for 2 hours.
4. Transfer rack(s) with slides into a tray with acetone and wash on shaker for 10-15 minutes. Repeat this step 2 more times by transferring rack(s) into trays with fresh acetone.
5. After last wash quickly transfer racks with slides into vacuum oven and dry in vacuum at room temperature for 20-30 minutes. Vacuum should be applied as fast as possible.
- 10 6. Dispose Pyridine-DMF mixture and acetone washes into flammable wastes container.
7. Transfer slides for storage into dry cabinets. Make sure the desiccant in the dry cabinet is good (blue in color).

Example 3. Printing of oligonucleotides.

- 15 Oligonucleotides used in this experiment were dissolved in 0.1 M NaOH at 100 nanogrammm per microliter and printed on PDITC modified glass surface. Amount of DNA deposited was about 5 ng per spot. After printing slides were baked at 80 °C for 2 hours and then UV crosslinked (254 nm UV lamp) for 1 min.

20 **Example 4. Preparation of Array**

Using the above protocol, an array having the characteristics of Table 1 was prepared. Each of the probe oligonucleotides was prepared using an automated nucleic acid synthesizer.

TABLE I

Array Position	Probe	Probe Sequence
A1	s64_2	AC CTAGAAAGCT ATTGAGCTG GATCGGTGCC TGTGATCGTG AGCGTTCTCT TGAAGAAATTT CGGACATCTC TGCCAAAGTC TTGTGACCTG TAGCTGCCA (NO:03)
A2	s64_2_90	AGAAAGCTATTGAGCTGGATCGGTCCCTCTGATCGTACGCTCTCTTGAAGAAATTCGGACATCTCTGCCAAAGTCTTGTGACCTGTA (SEQ ID NO:04)
A3	s64_2_80	AGCTATTGAGCTGGATCGGTCCCTCTGATCGTACGCTCTCTTGAAGAAATTCGGACATCTCTGCCAAAGTCTTGTGTA (SEQ ID NO:05)
A4	s64_2_70	ATTGAGCTGGATCGGTCCCTCTGATCGTACGCTCTCTTGAAGAAATTCGGACATCTCTGCCAAAGTA (SEQ ID NO:06)
B1	s64_2_60	AGCTGGATCGGTCCCTCTGATCGTACGCTCTCTTGAAGAAATTCGGACATCTCTGCCA (SEQ ID NO:07)
B2	s64_2_50	ATCCGCTCCTCTGATCGTACGCTCTCTTGAAGAAATTCGGACATCTA (SEQ ID NO:08)
C1	s26_2	AAACCCAGGA AAATACCAA TCCAGATTTT TTTGAGATC TGGAACTTT CAGATGACT CCTTTTAGTG CTATTGGTTT GGAGCTGTGG TCCATGACCTA (NO:09)
C2	s26_2_90	AGGAAATACCAATCCAGATTTCTTTGAAGATCTGGAACTTTTCAAGATGACTCCTTTTAGTGCTATTGGTTTGGAGCTGTGGTCCATA (ID NO:10)
C3	s26_2_80	AAATACCAATCCAGATTTCTTTGAAGATCTGGAACTTTTCAAGATGACTCCTTTTAGTGCTATTGGTTTGGAGCTGTGGA (SEQ ID NO:11)
C4	s26_2_70	AAATCCAGATTTCTTTGAAGATCTGGAACTTTTCAAGATGACTCCTTTTAGTGCTATTGGTTTGGAGCA (SEQ ID NO:12)
D1	s26_2_60	ACAGATTTCTTTGAAGATCTGGAACTTTTCAAGATGACTCCTTTTAGTGCTATTGGTTTA (SEQ ID NO:13)
D2	s26_2_50	ATTCTTTGAAGATCTGGAACTTTTCAAGATGACTCCTTTTAGTGCTATTA (SEQ ID NO:14)
A5 and E5	c370_2	AGGGTC AGCTGATCTA CGAGTCTGCC ATCACCTGTG AGTACTGGA TGAAGCATAC CCAGGGAAGA AGCTGTGTGCC GGATGACCCC TATGAGAAAG CTTGCA (NO:15)
A6 and E6	c370_2_90	AGCTGATCTAGAGTCTGCCATCTGCTGTGAGTACCTGATGAGCATACCCAGGGAAGAGCTGTTCGCGATGACCCCTATGAGAAA (SEQ ID NO:16)
A7 and E7	c370_2_80	ATCTAGAGTCTGCCATCTGCTGTGAGTACCTGATGAGCATACCCAGGGAAGAGCTGTTCGCGATGACCCCTATA (SEQ ID NO:17)
A8 and E8	c370_2_70	ACGAGTCTGCCATCACCTGTGAGTACCTGATGAGCATACCCAGGGAAGAGCTGTTCGCGATGAGCA (SEQ ID NO:18)
B5 and F5	c370_2_60	ACTGCCATCACCTGTGAGTACCTGATGAGCATACCCAGGGAAGAGCTGTTCGCGNA (SEQ ID NO:19)
B6 and F6	c370_2_50	ATCACCTGTGAGTACCTGATGAGCATACCCAGGGAAGAGCTGTGTA (SEQ ID NO:20)
G1	s91_3	AGGCCCCAAT GGCCTGGAAAT CTGCGCTATT TAGGCATTCT ACTCAGAAA ACCTTAAAAA TTCACAAATG TGTCAAGA GCCTTGATGT GGAAACCGATA (ID NO:21)
G2	s91_3_90	ACAAATGGCTGGAATCTGCTATTAGGCATTCTACTCAGAAAACCTTAAAAATTCACAAATGTCAGAGAGCCCTTGATGTGNA (SEQ ID NO:22)
G3	s91_3_80	AGGCTGGAATCTGCGCTATTAGGCATTCTACTCAGAAAACCTTAAAAATTCACAAATGTCAGAGAGCCCTTGATA (SEQ ID NO:23)
G4	s91_3_70	AGAAATCTGCTATTAGGCATTCTACTCAGAAAACCTTAAAAATTCACAAATGTCAGAGAGCCCA (SEQ ID NO:24)
H1	s91_3_60	ACTCGCTATTAGGCATTCTACTCAGAAAACCTTAAAAATTCACAAATGTCAGAAA (SEQ ID NO:25)
H2	s91_3_50	ACTATTAGGCATTCTACTCAGAAAACCTTAAAAATTCACAAATGTGTA (SEQ ID NO:26)
E1	s97_4	ATAGAGGGG TGAAGCCAG CTGCTCATGA ACGAGTTTGA GTCAGCCNAG GGTGACTTTG AGAAGTGT GGAAGTAAAC CCCAGATA AGGCTGCAAGA (NO:27)
E2	s97_4_90	AGGGTGAAGCCAGCTGCTCATGACGAGTTTGAGTCAGCCAGGGTGACTTTGAGAAAGTCTGGAGTAAACCCCAAGATPAGGCA (SEQ ID NO:28)
E3	s97_4_80	AGAGCCAGCTGCTCATGACGAGTTTGAGTCAGCCAGGGTGACTTTGAGAAAGTCTGGAGTAAACCCCAAGATA (SEQ ID NO:29)
E4	s97_4_70	ACCAGCTGCTCATGACGAGTTTGAGTCAGCCAGGGTGACTTTGAGAAAGTCTGGAGTAAACCCCA (SEQ ID NO:30)
F1	s97_4_60	ATGCTCATGAACGAGTTTGAGTCAGCCAGGGTGACTTTGAGAAAGTCTGGAGTAAAC (SEQ ID NO:31)
F2	s97_4_50	AATGAACGAGTTTGAGTCAGCCAGGGTGACTTTGAGAAAGTCTGGAAA (SEQ ID NO:32)

C5 s74_3 ATATGT AACTGAGAA GGTGACAGTC CTTTGGGTGA CCATGTGGGT TCTCTGTCAG AGAATTAGC AGCAGTCGTC AATAACCTAA ATACTGGGCA AGTGTA (NO:33)
C6 s74_3_90 ARACTGAGRAGGTGACAGTCCTTTGGGTGACCATGTGGGTCTCTGTGAGAGAAATTAGCAGCAGTCGTCRAATAACCTRAATACTGGGA (SEQ ID NO:34)
C7 s74_3_80 AAAGAGGTGACAGTCCTTTGGGTGACCATGTGGGTCTCTGTGAGAGAAATTAGCAGCAGTCGTCRAATAACCTRAATAA (SEQ ID NO:35)
C8 s74_3_70 AAGTGACAGTCCTTTGGGTGACCATGTGGGTCTCTGTGAGAGAAATTAGCAGCAGTCGTCRAATAACCTA (SEQ ID NO:36)
D5 s74_3_60 ACAGTCCTTTGGGTGACCATGTGGGTCTCTGTGAGAGAAATTAGCAGCAGTCGTCRAATA (SEQ ID NO:37)
D6 s74_3_50 ACTTGGGTGACCATGTGGGTCTCTGTGAGAGAAATTAGCAGCAGTCGA (SEQ ID NO:38)

TABLE 1 (CONT)

Example 5 - Hybridization ³³P-labeled cDNA Target with oligo glass ARRAY

1. Prepare a solution of 6×SSC buffer containing 0.1% SDS.
2. Place glass slide with printed oligo DNA in a hybridization chamber and add 2 ml of the solution prepared in step 1.
- 5 3. Prehybridize for 30 min at 60°C.
4. Mix labeled cDNA probe (Example 1, about 200 μl, total about 2-5×10⁶ cpm) with 1/10th of the total volume (about 22 μl) of 10×denaturing solution (1 M NaOH, 10 mM EDTA) and incubate at 65°C for 20 min. Then add 5 μl (1 μg/μl) of human Cot-1 DNA, and equal volume (about 225 μl) of 2× Neutralizing solution (1M NaHPO₄, pH 7.0) and continue
10 incubating at 65°C for 10 min.
5. Add the mixture prepared in Step 4 to the 2 ml of solution prepared in Step 1. Make sure that the two solutions are mixed together thoroughly.
6. Pour out the prehybridization solution and discard. Replace with the solution prepared in Step 5.
- 15 7. Hybridize overnight at 60°C.
8. Carefully remove the hybridization solution and discard in an appropriate container. Place the glass slides in a washing chamber with 20 ml of Wash Solution 1 (2×SSC, 0.1% SDS). Wash the ARRAY for 10 min with continuous agitation at room temperature. Repeat this step four times.
- 20 9. Perform one additional 10-min wash in 20 ml of Wash Solution 2 (0.1× SSC, 0.1% SDS) with continuous agitation at room temperature.
10. Using forceps, remove the cDNA ARRAY from the container and shake excess the wash solution. Rinse with distilled water and let the array dry on air.
11. Expose the glass slide Array to X-ray film at -70°C with an intensifying screen.
- 25 Alternatively, use a phosphorimager (Molecular Dynamics).

Example 6. Assay for Hybridization Efficiency

- Using the arrays and above protocols, the hybridization efficiency of each probe of different length on the array described in Example 4 was assayed using ³²P labeled target complementary for
30 each of the probes. The results of this assay are provided in Fig. 1. The results demonstrate that a significant increase in hybridization efficiency is achieved with oligonucleotide probes having a length greater than 50 nt.

It is evident from the above discussion that the subject arrays provide for a significant advance in the field. The subject invention provides for arrays of probes in which all of the probes on the array have substantially the same level of high hybridization efficiency for their respective targets and exhibit a minimal level of non-specific hybridization. As such, the subject arrays
5 eliminate the need for using multiple probe sequences for each target of interest or using mismatch control probes for each target, which is at least desired if not required with other array formats. In addition, the arrays are readily fabricated using non PCR based protocols, where the fabrication process is suitable for use in high throughput manufacturing. As such, the subject arrays combine the benefits of high throughput manufacturability of short oligonucleotide arrays with the benefits
10 of high specificity observed in cDNA arrays. Accordingly, the subject invention represents a significant contribution to the art.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior
15 to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from
20 the spirit or scope of the appended claims.

WHAT IS CLAIMED IS:

1. An array comprising at least one pattern of probe oligonucleotide spots stably associated with the surface of a solid support, wherein each probe oligonucleotide spot of said pattern
5 corresponds to a target nucleic acid and comprises an oligonucleotide probe composition made up of long oligonucleotide probes that range in length from about 50 to 120 nt.
2. The array according to Claim 1, wherein two or more different target nucleic acids are represented in said pattern.
- 10 3. The array according to Claims 1 or 2, wherein each probe oligonucleotide spot in said pattern corresponds to a different target nucleic acid.
4. The array according to Claims 1, 2 or 3, wherein each long oligonucleotide probe on said
15 array has a high hybridization efficiency for its respective target.
5. The array according to any of the preceding claims, wherein each long oligonucleotide of said array has a low propensity for non-specific hybridization.
- 20 6. The array according to any of the preceding claims, wherein each of said probe long oligonucleotides of said array exhibit substantially the same high hybridization efficiency for their respective targets.
7. The array according to any of the preceding claims, wherein said long oligonucleotide
25 probes are covalently attached to said surface of said substrate.
8. The array according to Claim 7, wherein said each of said long oligonucleotide probes is cross-linked to the surface of said support at at least one site.
- 30 9. The array according to Claim 7, wherein each of said oligonucleotide probes is cross-linked to the surface of said support at at least two sites.
10. The array according to any of the preceding claims, wherein the density of spots on said array does not exceed about 1000/cm².

11. The array according to Claim 10, wherein the density of spots on said array does not exceed about 400/cm².
- 5 12. The array according to any of the preceding claims, wherein the number of spots on said array ranges from about 50 to 50,000.
13. The array according to Claim 12, wherein the number of spots on said array ranges from about 50 to 10,000.
- 10 14. The array according to any of the preceding claims, wherein the length of each long oligonucleotide ranges from about 60 to 100 nt.
15. The array according to any of the preceding claims, wherein ten or more different target
15 nucleic acids are represented in said pattern.
16. The array according to any of the preceding claims, wherein the length of each of said unique oligonucleotides ranges from about 65 to 90 nucleotides.
- 20 17. A method of preparing an array according to any of Claims 1 to 16, said method comprising:
generating said long oligonucleotide probes; and
stably associating said long oligonucleotide probes on the surface of said solid support in a
manner sufficient to produce said array.
- 25 18. A hybridization assay comprising the steps of:
contacting at least one labeled target nucleic acid sample with an array according to any of
Claims 1 to 16 under conditions sufficient to produce a hybridization pattern; and
detecting said hybridization pattern.
- 30 19. The method according to Claim 18, where said method further comprises:
generating a second hybridization pattern; and
comparing said hybridization patterns.

20. A kit for use in a hybridization assay, said kit comprising:
an array according to any of Claims 1 to 16.

1/1

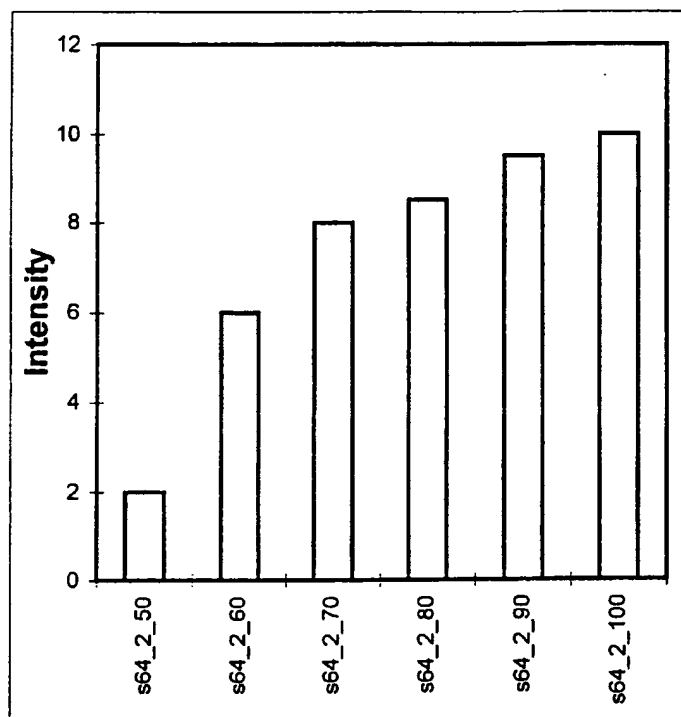


Figure 1

SEQUENCE LISTING

<110> Chenchik, Alex
Munishkin, Alexander
Simonenko, Peter

<120> Long Oligonucleotide Arrays

<130> CLON-015WO

<150> 09/440,829

<151> 1999-11-15

<160> 38

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 293

<212> DNA

<213> Artificial Sequence

<220>

<223> control oligonucleotide

<400> 1

ggccaggata ccaaagcctt acaggacttc ctcctcagtg tgcagatgtg cccaggtaat	60
cgagacactt actttcacct gcttcagact ctgaagaggc tagatcggag ggatgaggcc	120
actgcactct ggtggaggct ggaggcccaa actaaggggt cacatgaaga tgctctgtgg	180
tctctcccc tgtacctaga aagctatttg agctggatcc gtccctctga tcgtgacgcc	240
ttccttgaag aatttcggac atctctgcca aagtcttggt acctgtagct gcc	293

<210> 2

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> primer

<400> 2

cggccaggat accaaagcct tacag	25
-----------------------------	----

<210> 3

<211> 101

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 3

acctagaaaag ctatttgagc tggatccgtc cctctgatcg tgacgccttc cttgaagaat	60
ttcggacatc tctgccaaag tcttgtgacc tgtagctgcc a	101

<210> 4

<211> 90

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 4

agaaagctat ttgagctgga tccgtccctc tgatcgtgac gccttccttg aagaatttcg 60
gacatctctg ccaaagtctt gtgacctgta 90

<210> 5

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 5

agctatttga gctggatccg tccctctgat cgtgacgcct tccttgaaga atttcggaca 60
tctctgccaa agtcttgtga 80

<210> 6

<211> 70

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 6

atttgagctg gatccgtccc tctgatcgtg acgccttcct tgaagaattt cggacatctc 60
tgccaaagta 70

<210> 7

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 7

agctggatcc gtccctctga tcgtgacgcc ttccttgaag aatttcggac atctctgcca 60

<210> 8

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 8

aatccgtccc tctgatcgtg acgccttcct tgaagaattt cggacatcta 50

<210> 9

<211> 101

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 9

aaacccagga aaataccaaa tccagatttc tttgaagatc tggaaccttt cagaatgact 60
ccttttagtg ctattggttt ggagctgtgg tccatgacct a 101

<210> 10
<211> 90
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 10
aggaaaaatac caaatccaga tttctttgaa gatctggaac ctttcagaat gactcctttt 60
agtgctattg gtttgagct gtggtccata 90

<210> 11
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 11
aataccaaat ccagatttct ttgaagatct ggaacctttc agaatgactc cttttagtgc 60
tattggttg gagctgtgga 80

<210> 12
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 12
aaaatccaga tttctttgaa gatctggaac ctttcagaat gactcctttt agtgctattg 60
gtttggagca 70

<210> 13
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 13
acagatttct ttgaagatct ggaacctttc agaatgactc cttttagtgc tattggttta 60

<210> 14
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 14
attctttgaa gatctggaac ctttcagaat gactcctttt agtgctatta 50

<210> 15
<211> 102
<212> DNA
<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 15

agggtcagct gatctacgag tctgccatca cctgtgagta cctggatgaa gcatacccag 60
ggaagaagct gttgccggat gacccctatg agaaagcttg ca 102

<210> 16

<211> 90

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 16

aagctgatct acgagtcctgc catcacctgt gagtacctgg atgaagcata cccaggggaag 60
aagctgttgc cggatgaccc ctatgagaaa 90

<210> 17

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 17

aatctacgag tctgccatca cctgtgagta cctggatgaa gcatacccag ggaagaagct 60
gttgccggat gacccctata 80

<210> 18

<211> 70

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 18

acgagtcctgc catcacctgt gagtacctgg atgaagcata cccaggggaag aagctgttgc 60
cggatgacca 70

<210> 19

<211> 60

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 19

actgccatca cctgtgagta cctggatgaa gcatacccag ggaagaagct gttgccggaa 60

<210> 20

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 20

aatcacctgt gagtacctgg atgaagcata cccaggggaag aagctgttga 50

<210> 21
<211> 102
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 21
aggccccaaa tggctggaaa tctcgcttat ttaggcattc tactcagaaa aaccttaaaa 60
attcacaaat gtgtcagaag agccttgatg tggaaaccga ta 102

<210> 22
<211> 90
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 22
acaaatggct ggaaatctcg cctatttagg cattctactc agaaaaacct taaaaattca 60
caaatgtgtc agaagagcct tgatgtggaa 90

<210> 23
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 23
aggctggaaa tctcgcttat ttaggcattc tactcagaaa aaccttaaaa attcacaaat 60
gtgtcagaag agccttgata 80

<210> 24
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 24
agaaatctcg cctatttagg cattctactc agaaaaacct taaaaattca caaatgtgtc 60
agaagagcca 70

<210> 25
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 25
actcgcttat ttaggcattc tactcagaaa aaccttaaaa attcacaaat gtgtcagaaa 60

<210> 26
<211> 50

<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 26
actatttagg cattctactc agaaaaacct taaaaattca caaatgtgta 50

<210> 27
<211> 101
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 27
ataggagggg tgaagcccag ctgctcatga acgagtttga gtcagccaag ggtgactttg 60
agaaagtgtc ggaagtaaac cccagaata aggtgtcaag a 101

<210> 28
<211> 90
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 28
aggggtgaag cccagctgct catgaacgag tttgagtcag ccaaggggtga ctttgagaaa 60
gtgctggaag taaacccccca gaataaggca 90

<210> 29
<211> 80
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 29
agaagcccag ctgctcatga acgagtttga gtcagccaag ggtgactttg agaaagtgtc 60
ggaagtaaac cccagaata 80

<210> 30
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 30
accagctgct catgaacgag tttgagtcag ccaaggggtga ctttgagaaa gtgctggaag 60
taaacccccca 70

<210> 31
<211> 60
<212> DNA
<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 31
atgctcatga acgagtttga gtcagccaag ggtgactttg agaaagtgct ggaagtaaaa 60

<210> 32

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 32
aatgaacgag tttagtcag ccaaggggta ctttgagaaa gtgctggaaa 50

<210> 33

<211> 102

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 33
atatgtaact gaagaagggtg acagtccttt gggtgaccat gtgggttctc tgtcagagaa 60
attagcagca gtcgtcaata acctaaatac tgggcaagtg ta 102

<210> 34

<211> 90

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 34
aaactgaaga aggtgacagt cctttgggtg accatgtggg ttctctgtca gagaaattag 60
cagcagtcgt caataaccta aatactggga 90

<210> 35

<211> 80

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 35
aaagaagggtg acagtccttt gggtgaccat gtgggttctc tgtcagagaa attagcagca 60
gtcgtcaata acctaaataa 80

<210> 36

<211> 70

<212> DNA

<213> Artificial Sequence

<220>

<223> synthetic oligonucleotide

<400> 36
aagtgcagct cctttgggtg accatgtggg ttctctgtca gagaaattag cagcagtcgt 60
caataaccta 70

<210> 37
<211> 60
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 37
acagtccttt gggtgaccat gtgggttctc tgtcagagaa attagcagca gtcgtcaata 60

<210> 38
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 38
actttgggtg accatgtggg ttctctgtca gagaaattag cagcagtcga 50

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36682 A3

- (51) International Patent Classification⁷: **C12Q 1/68**
- (21) International Application Number: **PCT/US00/31562**
- (22) International Filing Date:
15 November 2000 (15.11.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/440,829 15 November 1999 (15.11.1999) US
- (71) Applicant: **CLONTECH LABORATORIES, INC.**
[US/US]; 1020 East Meadow Circle, Palo Alto, CA 94303 (US).
- (72) Inventors: **CHENCHIK, Alex**; 670 San Antonio Road #30, Palo Alto, CA 94306 (US). **MUNISHKIN, Alexander**; 109 Castillion Terrace, Santa Cruz, CA 95060 (US). **SIMONENKO, Peter**; 712 Sierra Vista, Mountain View, CA 94043 (US).
- (74) Agent: **FIELD, Bret, E.**; Bozicevic, Field & Francis LLP, Suite 200, 200 Middlefield Road, Menlo Park, CA 94025 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- (88) Date of publication of the international search report:
17 January 2002
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 01/36682 A3

(54) Title: **LONG OLIGONUCLEOTIDE ARRAYS**

(57) Abstract: Long oligonucleotide arrays, as well as methods for their preparation and use in hybridization assays, are provided. The subject arrays are characterized in that at least a portion of the probes of the array, and usually all of the probes of the array, are long oligonucleotides, e.g. oligonucleotides having a length of from about 50 to 120 nt. Each long oligonucleotide probe on the array is preferably chosen to exhibit substantially the same high target binding efficiency and substantially the same low non-specific binding under conditions in which the array is employed. The subject arrays find use in a number of different applications, e.g. differential gene expression analysis.

INTERNATIONAL SEARCH REPORT

Intern. Application No

PCT/US 00/31562

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS, CHEM ABS Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GB 2 318 791 A (ZENECA LTD) 6 May 1998 (1998-05-06) claims 1-10	1-20
X	US 5 807 522 A (BROWN PATRICK O ET AL) 15 September 1998 (1998-09-15) column 4, line 16 - line 26; claims 1-7	1-20
X	US 5 830 645 A (ALBERTSON DONNA ET AL) 3 November 1998 (1998-11-03) column 13 -column 14; claim 1	1-20

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *B* document member of the same patent family

Date of the actual completion of the international search

9 August 2001

Date of mailing of the international search report

16/08/2001

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Gabriels, J

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern. Application No

PCT/US 00/31562

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
GB 2318791 A	06-05-1998	EP 0935671 A WO 9818961 A JP 2001502909 T	18-08-1999 07-05-1998 06-03-2001
US 5807522 A	15-09-1998	AT 180570 T AU 709276 B AU 2862995 A CA 2192095 A DE 69509925 D DE 69509925 T DK 804731 T EP 0804731 A EP 0913485 A ES 2134481 T GR 3030430 T JP 10503841 T US 6110426 A WO 9535505 A	15-06-1999 26-08-1999 15-01-1996 28-12-1995 01-07-1999 09-12-1999 08-11-1999 05-11-1997 06-05-1999 01-10-1999 30-09-1999 07-04-1998 29-08-2000 28-12-1995
US 5830645 A	03-11-1998	EP 0800587 A JP 11510681 T WO 9617958 A	15-10-1997 21-09-1999 13-06-1996